Assessing Individual Differences in a Phishing Detection Task

Christopher B. Mayhorn[a], Allaire K. Welk[a], Olga A. Zielinska [a], and Emerson Murphy-Hill [b]

[a]*Department of Psychology, North Carolina State University, Raleigh, North Carolina, USA*
[b]*Department of Computer Science, North Carolina State University, Raleigh, North Carolina, USA*

1. Introduction

Some authors suggest that regardless of how good security technology is, it is the “people problem” that must be overcome for successful cybersecurity (West, Mayhorn, Hardee, & Mendel, 2009). While security threats to the average computer user might take a variety of forms such as viruses or worms delivered via nefarious websites or USB drives, identity theft tactics such as phishing are becoming increasingly problematic and common. Phishing is a technology-based, social engineering tactic where attackers attempt to appear as authorized sources to target individuals and obtain personal and/or sensitive information. The current research aims to explore how individuals differ in phishing susceptibility within the context of a real world email-related decision making task.

2. Method

Researchers recruited fifty-three undergraduate students attending North Carolina State University from an Introductory Psychology course and an Introductory Computer Science course. Prior to attending an in-person experimental appointment**,** an online survey was administered where participants were asked to rate the extent to which various statements represented their disposition and behaviors**.** The survey used a number of established questionnaire instruments to assess the following information: basic demographics, personality characteristics, dispositional trust, impulsivity, and web/computer based behavior, beliefs, and previously experienced phishing consequences.

2.1 Materials

Participants enrolled in this study through North Carolina State University’s online experiment scheduling system. Prior to attending an in-person experimental appointment, an online survey was administered where participants were asked to rate the extent to which various statements represented their disposition and behaviors. The survey used a number of established questionnaire instruments to assess the following information: basic demographics, personality characteristics (Gosling, Rentfrow, & Swann, 2003), dispositional trust (Merritt & Ilgen, 2008), impulsivity (Nyeste & Mayhorn, 2009), and web/computer based behavior, beliefs, and previously experienced phishing consequences (Eveland, Shah, & Kwak, 2003; Yoshioka, Washizaki, & Maruyama, 2008; Kelley, Hong, Mayhorn, Murphy-Hill, 2012). Participants responded to trust and impulsivity items using a 5-point Likert scale (1=Very Inaccurate, 5=Very Accurate) and personality items utilizing a 7-point Likert scale (1=Strongly Disagree, 7=Strongly Agree). Additionally, participants indicated whether or not they perform various behavioral actions with a dichotomous (yes/no) response. Survey information was generated, presented, and stored through Qualtrics Survey Software. Once participants arrived for the experiment appointment, they demonstrated that they had normal/corrected-to-normal vision and could accurately perceive the presented computer-based information by completing a Snellen eye examination for far vision.

Researchers implemented a computer-based email categorization/decision making task based on a previously constructed role-play exercise, proven to be internally and externally valid by Sheng, Holbrook, Kumaraguru, Cranor, Downs (2010). The benefit of utilizing a roleplay task, rather than having students behave as themselves, is that it allows researchers to study phishing and human behavior without conducting an actual phishing attack (Sheng, Holbrook, Kumaraguru, Cranor, Downs, 2010).

Participants were informed that they were to adopt the role of William Smith/Mary Johnson, a 50-year-old male/female who works at North Carolina State University as a staff member. This contextual information replicates the details presented in Sheng, Holbrook, Kumaraguru, Cranor, & Downs (2010) with the exception of the University and certain content included in the emails. The deviations implemented aimed to present contemporary emails that fit the context provided in the role-play scenario. The emails presented varied in content (i.e. banking, scholarships, etc.), and were included to assess behavior with a wide range of message types; a manipulation consistent with that of Sheng, Holbrook, Kumaraguru, Cranor, & Downs (2010).

This task was presented through a Firefox browser. An experiment-specific Gmail email account was created and utilized to simulate a real-world email task. Participants' actions and responses were recorded using the screen capture function of QuickTime Player. All data analysis was conducted using Microsoft Excel and SPSS statistical software.

2.2 Procedure

Upon registering for this experiment online, participants were assigned a unique participant identification number and sent an online survey. Informed consent was obtained prior to completion of the survey. This survey collected participants' self-report data, and was completed in advance of each in-person experiment appointment.

When participants arrived for their in-person experimental appointment, the experimenter confirmed that they had completed the aforementioned survey. Following confirmation, participants performed a Snellen far vision test; all participants illustrated normal or corrected-to-normal vision.

Participants were then directed to an experimental computer to perform a two-part email task. Within part one, participants performed the aforementioned role-play email categorization task; they were instructed to adopt the role of William Smith/Mary Johnson, a 50-year-old man/woman, respectively.

The experimenter read the participant a brief narrative that outlined details regarding their assumed role, specifically including characteristics, occupation, career objectives, and their daily email routine. Participants were directed to replicate this daily email routine. First, they were to log into the email account using account credentials provided by the experimenter and send an email to their mother explaining what they had done the day before. They were informed that this did not need to be a lengthy task; they were instructed to send the email within 5 minutes. This initial email task was included to simulate a genuine email task and authenticate this email process for the participant, specifically for the purpose of inspiring actions representative of their typical email behavior.

Next, participants performed the primary email task. They were instructed to check the inbox of the provided email account, access each new email, read it, and evaluate each message as they normally would (e.g. if they normally skim the message, then they should do so within this task). Following this evaluation process, participants were to select the most appropriate action by deciding to archive, flag as important, or delete each email. Fourteen emails were presented to each participant. This data was captured and recorded through the screen capture function of QuickTime player. Responses were later coded as correct or incorrect for both types of emails (legitimate or phishing). A response for a legitimate email was coded as correct if the participant flagged it as important or archived it; whereas a response for a phishing email was coded as correct if the participant deleted it. A response for a SPAM email was coded as correct if the participant archived or deleted it. This coding scheme facilitated the generation of two primary performance measurements that are discussed within the results section of this paper. Upon completion of the email task, participants were provided with a debriefing form that discussed the objectives and methods of the study.

3. Results

For each participant, an overall accuracy score was computed; this measure was calculated by dividing the total number of correctly identified emails, including both legitimate emails and phishing attempts, by the total number of all presented emails. Additionally, a phishing accuracy score was calculated; this measure was derived by dividing the number of correctly identified phishing emails by the total number of exclusively phishing emails presented. A high score on the overall accuracy measure indicates high accuracy in identifying both legitimate and phishing emails, whereas a high score on the phishing accuracy measure demonstrates high performance in identifying exclusively phishing attempts; for both of these accuracy measures, the highest score possible was 100% accurate (descriptive statistics for these measures are presented in Table 1).

Table 1. *Descriptive statistics for accuracy measures.*

	M	*SD*	Range	Minimum	Maximum
Overall Accuracy	0.64	0.12	0.57	.43	1.00
Phishing Accuracy	0.46	0.21	0.83	.17	1.00

Based on the pattern of correlations, a hierarchal regression analysis was conducted to determine which variables best predicted phishing detection accuracy (see Table 2). The first model entered into the regression equation included the aforementioned personality and impulsivity predictors. Results indicate that these variables were significant unique predictors of phishing detection accuracy, *F*(5, 46) = 4.52, *p* = .002, R^2 = .33. The second model inputted to the regression added the trust/distrust predictors; results indicate that these variables were also significant predictors, *F*(8, 43) = 3.81, *p* = .002, R^2 = .42, but did not significantly contribute to the model, *F* change= 2.10, *p* = .116, R^2 change = .01 Lastly, the third model inputted added the behavioral independent variables; results indicate that these variables were significant predictors, *F*(11, 40) = 4.08, *p* < .001, R^2 = .53 and significantly contributed to the model, *F* change = 3.42, *p* = .032, R^2 change = .11.

Table 2. *Hierarchal regression analyses predicting phishing accuracy.*

Model	β	R^2	$R^2Δ$	*F*	*p*
Impulsivity/Personality **Items**: Model 1		.33	.33	4.52	.002
Extraversion	-.11				
Anxiety	-.25				
Reservation	.25				
Calmness	.04				
Ability to keep emotions under control	.14				
Trust/Distrust **Items**: Model 2		.42	.09	3.81	.002
Extraversion	.07				
Anxiety	-.33				
Reservation	.29				
Calmness	.04				
Ability to keep emotions under control	.14				
Trust what people say	-.14				
Believe others have good intentions	.01				
General distrust	.26				
Behavioral Measures **Items**: Model 3		.53	.11	4.08	< .001
Extraversion	.27				
Anxiety	-.25				
Reservation	.15				
Calmness	.68				
Ability to keep emotions under control	.39				
Trust what people say	-.10				
Believe others have good intentions	-.07				
General distrust	.16				
Lost money, was never reimbursed	.32				
Completely read phishing message	.02				

4. Discussion

Results from this study provide further evidence that there are individual differences in phishing susceptibility; moreover, both dispositional and behavioral factors can explain this variation in phishing detection ability across individuals. Additionally, there are different factors that predict how individuals discriminate legitimate emails from phishing attempts.

In terms of managing both phishing and legitimate emails, individuals' accuracy was affected by: reservation, the ability to keep emotions under control, distrust, a belief that others are essentially evil, losing money without being reimbursed, and a belief that one may receive a legitimate request to confirm account information via email. These results support the previous findings of Wright and Marett (2012) in that individuals who are suspicious of others and exhibit a general distrust toward people are less susceptible to phishing attacks. Additionally, the current findings are consistent with some of those reported by Pattinson and Jerram (2012): low impulsivity was related to elevated performance. Taken together, these results suggest that personality characteristics that support reserved behavior, low impulsivity, and distrust decreased phishing susceptibility within an email-based decision making task.

The present study also provides two behavioral/consequence factors that were related to phishing susceptibility: experiencing a monetary loss without eventual reimbursement and a belief that one may receive a legitimate request to confirm account information via email. The latter finding: a belief that one may receive a legitimate request to confirm account information via email was positively related to overall accuracy; this finding is seemingly counterintuitive. At face value, one might infer that this belief could lead to increased phishing susceptibility. Future research should aim to replicate and examine this effect more closely.

Due to limited prior research on susceptibility with exclusively phishing emails, we cannot make comparisons between the current study's findings and those of previously conducted studies. However, within the present study identifying exclusively phishing emails was influenced by the following predictors: extraversion, anxiety, reservation, calmness, keeping emotions under control, distrust, completely reading email messages, and losing money without being reimbursed. As indicated through the results of hierarchal regression, the best predictors for phishing accuracy were personality and impulsivity items; these results suggest that personality and impulsivity predictors are highly related to overall email categorization performance and phishing susceptibility. These data provide applications, in terms of establishing human-centered anti-phishing countermeasures. This knowledge can aid in the generation of individualized anti-phishing training programs and/or anti-phishing technologies personalized to target specific dispositional characteristics. For example: if individuals who are more impulsive than others can be identified, these individuals can then be targeted for explicit training procedures or specialized anti-phishing technologies, thereby reducing phishing susceptibility. The feasibility of this manipulation should be examined in future studies.

References

Eveland, W. P., Shah, D. V., & Kwak, N. (2003). Assessing causality in the cognitive mediation model: A panel study of motivations, information processing, and learning during campaign 2000. *Communication Research, 30*(4), pp. 359-386.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the big five personality domains. *Journal of Research in personality, 37*(6), pp. 504-528.

Kelley, C. M., Hong, K. W., Mayhorn, C. B., & Murphy-Hill, E. (2012). Something smells phishy: Exploring definitions, consequences, and reactions to phishing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56(1), pp. 2108-2112. SAGE Publications.

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*(2), pp. 194- 210.

Nyeste, P. G., & Mayhorn, C. B. (2009). Perceptions of cybersecurity: An exploratory analysis. *Proceedings of the 17th world congress of the international ergonomics association.* Beijing, China.

Pattinson, M., Jerram, C. (2012). Why do some people manage phishing e-mails better than others? *Information Management and Computer Security*, 20(1), pp. 18-28.

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. (2010). Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 373-382.

West, R., Mayhorn, C. B., Hardee, J., & Mendel, J. (2009). The Weakest Link: A Psychological Perspective on

Why. *Social and Human Elements of Information Security: Emerging Trends,* pp. 43-60.

Wright, R.T. & Marett, K. (2010). The influence of experiential and dispositional factors in phishing: an empirical investigation of the deceived. *Journal of Management Information Systems,* 27(1), pp. 273-303.

Yoshioka, N., Washizaki, H., & Maruyama, K. (2008). A survey on security patterns. *Progress in Informatics, 5*(5), pp. 35-47.